

# マイクロブログ記事における人物の呼称表現に関する研究

山口 裕太郎

人物について言及する際に、私たちは、人名以外に様々な呼称を使用している。例えば斎藤佑樹投手の場合、ファンは斎藤投手を「ハンカチ王子」や「佑ちゃん」といった呼称で呼ぶことが多い。前者は斎藤投手の振る舞いに由来するものであり、後者は斎藤投手の名前に由来する。また、前田敦子には「あっちゃん」、「顔面センター」といった呼称が存在し、それぞれが使用される文脈や話題は異なると考えられる。このように、呼称は、単なる人物の参照手段にとどまらず、対象となる人物との距離感や親愛の程度を表す表現となっていると思われる。

一方、身近な人物や話題に関する記事を気軽に投稿するコミュニケーション手段として、マイクロブログが注目を集めている。マイクロブログでは、ユーザは、現在生じている事柄に対する感覚的な意見を手軽に投稿するので、記事中には多くの呼称が頻出すると考えられる。また、Web 上には既に人に関する情報が溢れている現状もある。以上のことから、人物の呼称の特徴を明らかにすることは、人物に関する情報検索や、テキストマイニング技術を飛躍的に向上させると期待される。

本論文では、人名以外の参照のされ方を呼称と定義し、人物に付けられた様々な呼称が持つ特徴を明らかにすることを目的に、呼称抽出及び呼称分析手法を提案する。具体的には、検索エンジンと Wikipedia から人物の呼称を収集する。従来の検索エンジンと機械学習を用いる手法に加えて、本研究では、Wikipedia も併用することで呼称を網羅的に収集する。呼称を用いて書かれた記事の投稿数や、評価極性を分析する。

呼称候補の分類の評価実験の結果、品詞情報を素性に加えた提案手法では正例で適合率 0.83 の精度を示し品詞情報を使用しない場合に比べて精度が向上した。抽出した呼称を使用して、記事の投稿件数を分析したところ、呼称でも十分な投稿がなされていることを確認した。また、投稿件数のピークが人名と呼称で異なる部分が見られた。これは、出来事によっては呼称でしか投稿されないものがあるためだと考えられる。記事の評価極性は、呼称と人名では異なる値を示した。分析結果を踏まえた呼称の利用法の一つとして、人物の呼称を用いたマイクロブログ検索システムを考案した。

今後の課題として、呼称抽出の精度向上、評価極性分析の精緻化、システム改良が挙げられる。

(指導教員 佐藤 哲司)